



HAL
open science

Valoriser une bibliothèque numérique par des choix de référencement et diffusion. L'expérience de la Bibliothèque Sainte-Geneviève

Mathieu Andro, Marion Chaigne, Franck Smith

► To cite this version:

Mathieu Andro, Marion Chaigne, Franck Smith. Valoriser une bibliothèque numérique par des choix de référencement et diffusion. L'expérience de la Bibliothèque Sainte-Geneviève. Les Cahiers du numérique, 2012, 9 (3), pp.75-90. hal-01094573

HAL Id: hal-01094573

<https://hal.science/hal-01094573v1>

Submitted on 4 Mar 2015

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Valoriser une bibliothèque numérique à travers des choix de référencement et de diffusion : l'expérience de la Bibliothèque Sainte-Geneviève

Mathieu Andro, Marion Chaigne, Franck Smith
Bibliothèque Sainte-Geneviève

Abstract

This feedback is about how digital projects at the Sainte-Genevieve Library have widely considered the diffusion questions of dissemination, SEO, but also communication and enhancement of digital library projects.

Résumé

Cet article de retour d'expérience vise à montrer comment les projets de numérisation à la Bibliothèque Sainte-Geneviève se sont efforcés de prendre en compte la question de la diffusion, du référencement, mais aussi de la communication au service de la valorisation de sa bibliothèque numérique.

Introduction

La Bibliothèque Sainte-Geneviève est une des grandes bibliothèques françaises dont l'originalité tient à la fois à son double statut de bibliothèque publique et universitaire, à son architecture unique, à sa longue et riche histoire comme à ses collections importantes et pour partie patrimoniales¹. Cette contribution vise à présenter un retour d'expérience en rendant compte d'une démarche privilégiant les aspects techniques et de gestion d'un projet de bibliothèque numérique tel que développé par la Bibliothèque Sainte-Geneviève.

Depuis une quinzaine d'années que des projets de numérisation sont menés en bibliothèque, il est désormais établi que la chaîne de production numérique ne saurait dès l'abord se concevoir sans une définition préalable de la chaîne de diffusion. La sélection des documents à numériser ne procède plus en effet prioritairement d'un souci de conservation mais vise également et surtout à mettre le plus en valeur possible un patrimoine tant documentaire que culturel. Au-delà des premières expériences qui pouvaient impliquer le recours à des supports physiques (cédéroms ou dévédéroms) dans le seul espace physique des bibliothèques, l'accès distant, simplifié et enrichi aux contenus en mode image et en mode texte s'est désormais imposé comme une condition nécessaire de tout projet : la publication sur le web de leur patrimoine numérisé permet potentiellement

¹ Peyré Y. (2011). *La bibliothèque Sainte-Geneviève à travers les siècles*, Paris, Gallimard.

aux bibliothèques de toucher un public virtuel éparpillé, diversifié, et surtout plus nombreux que le public fréquentant les lieux.

Dans la littérature professionnelle, les conférences et les formations spécifiques, aborder la numérisation revient généralement à l'envisager sous l'angle de la politique documentaire (quels corpus sélectionner, numérisation de masse ou numérisation de niches etc. ?), de la production et du contrôle de production (production interne ou, le plus souvent, externalisée en totalité ou en partie ?), de la conservation et de l'archivage pérenne (quels supports de conservation, quelles métadonnées et selon quels formats ?), ou encore sous l'angle sémantique, d'étude des usages (quels indicateurs bibliométriques ?), de conséquences sur la société de l'information (quelle gestion de droits, quelles médiations etc. ?). Il demeure en revanche moins fréquent de s'intéresser à cette question des stratégies et des modalités de diffusion conçues comme un élément de valorisation des documents mis en ligne. Il s'agit pourtant d'un enjeu véritable dans un contexte de multiplication à grande échelle des projets de numérisation du patrimoine des bibliothèques. Il faut à la fois prendre en compte l'irruption de nouveaux acteurs extérieurs au monde de la culture ou de la recherche, associés ou non aux institutions culturelles ou universitaires traditionnelles, dans la production et la valorisation de ces fonds d'une part et l'évolution très rapide des usages du web avec en particulier des internautes aujourd'hui largement acquis aux grands moteurs de recherche d'autre part. Dans les deux cas, c'est la masse exponentielle de données disponibles, leur interrogation et leur identification rapides, leur croisement et leur exploitation personnalisés qui constituent le défi principal : quelle exposition sur le web pour les données produites et quelle influence sur les publics ciblés ? Quelle valeur ajoutée, notamment en terme de services associés de manière libre ou onéreuse, par rapport à la multitude de contenus gratuits ou payants, uniques ou redondants déjà offerts ? Ces interrogations ne relèvent pas seulement d'une stratégie de communication ou de marketing postérieure à la numérisation du patrimoine écrit. Elles peuvent avantageusement prendre corps avant même le lancement de la production afin d'orienter des choix techniques, logiciels ou documentaires en faveur d'une meilleure visibilité de ce patrimoine écrit numérisé et d'un élargissement des publics potentiels.

En ce sens, la valorisation des corpus numérisés mérite aussi de s'envisager avant la numérisation proprement dite et pas uniquement après : elle passe par des choix préalables concernant la diffusion, c'est-à-dire l'accès aux documents, et le référencement, c'est-à-dire la visibilité sur le web des données produites. La dissémination de l'information par la multiplication

des accès aux documents numérisés, le nombre et l'individualisation des services associés aux contenus peuvent aussi partiellement compenser selon les cas la modestie des moyens matériels et humains à disposition d'un établissement ou la taille réduite d'un corpus en ligne et créer les conditions d'une valorisation efficace. Le rapport coût/usage des projets s'en trouve amélioré à l'heure où les considérations économiques pèsent de tout leur poids dans la conduite des projets de numérisation, qu'ils soient à l'initiative d'acteurs aussi bien privés que publics.

Le référencement, un enjeu essentiel de la valorisation

Dans la mesure où il semble que de nombreux internautes utilisent pour leur collecte d'information des moteurs de recherche² et ne privilégient que très secondairement l'accès direct aux sites Internet des bibliothèques, des institutions culturelles ou universitaires - quand bien même ils sont parfois parvenus à un degré de notoriété certain auprès du grand public ou de communautés plus restreintes - il est primordial pour les bibliothèques numériques constituées d'être correctement référencées. Il importe donc que les documents mis en ligne qui composent ces bibliothèques numériques apparaissent en tête des résultats lorsqu'un lecteur en saisira le titre ou tout autre élément d'identification bibliographique significatif dans la barre d'un moteur de recherche.

Privilégier l'orientation web d'un projet de numérisation

La Bibliothèque Sainte-Geneviève a entrepris en 2011 de numériser un ensemble de documents provenant de ses trois fonds documentaires (fonds de la Réserve, fonds fenno-scandinave de la Bibliothèque nordique, fonds général) : 300 incunables, 250 récits de voyages nordiques et 360 ouvrages du XIXe siècle peu répandus dans le réseau des bibliothèques françaises. Cet ensemble de documents libres de droits traduit la richesse et l'originalité des collections anciennes de la bibliothèque, aussi bien dans leur dimension patrimoniale que dans leur dimension encyclopédique. Le fil directeur qui a présidé à la définition de ce premier projet de numérisation a toutefois davantage été celui de la rareté relative de ces documents originaux dans les grandes collections universitaires ou publiques françaises et, après vérification des principales plateformes institutionnelles ou

² De Rosa C., Cantrell J., Cellentani D., Hawk J., Jenkins L., Wilson A. (2005). *Perceptions of libraries and information resources. A report to the OCLC membership*. « 84% of users began an information search with search engine ». Les statistiques de consultation de la bibliothèque numérique Sainte-Geneviève (origine des visiteurs) confirme cette analyse.

commerciales françaises et étrangères, le fait qu'ils n'étaient pas encore disponibles en ligne. Dès l'étape de la sélection, ce projet a donc été résolument orienté vers le web : ces documents aujourd'hui diffusés en ligne ne se suffisent pas à eux-mêmes pour former un corpus unique susceptible de satisfaire à lui seul des besoins clairement identifiés de publics bien déterminés et acquérir de la sorte une reconnaissance immédiate corrélée à l'institution qui les a produits. Ce projet, ambitieux pour l'établissement mais modeste au regard des plans de numérisation de masse en cours ou achevés, se conçoit donc d'abord comme une contribution à la diffusion gratuite et la plus large possible du patrimoine scientifique et culturel à partir des collections qu'il possède, sans redondance et en écho à des collections déjà numérisées par ailleurs. Il s'agit en ce sens d'abonder une bibliothèque numérique virtuelle à l'échelle du web ; la définition des corpus et le choix du point d'accès relèvent davantage des besoins documentaires ressentis par l'internaute et de ses pratiques de recherche d'information en ligne que du bibliothécaire ou documentaliste prescripteur. L'enjeu réside pour la bibliothèque dans le référencement optimal de chaque unité documentaire et la multiplication des points d'accès à ces unités documentaires plutôt que dans une visibilité immédiate de l'institution à travers un corpus en ligne délimité et proposé principalement depuis son site.

Tenir compte des algorithmes de référencement des moteurs de recherche pour améliorer la visibilité de la bibliothèque numérique

Lorsqu'on interroge un moteur de recherche, les résultats obtenus sont classés par ordre de pertinence. Ce classement est lui-même déterminé par un algorithme appelé PageRank³ chez le moteur de recherche le plus connu et le plus utilisé, Google. Bien que cet algorithme soit en grande partie secret, il est fondé sur la mesure de la « popularité » d'un site, elle-même mesurée à partir du nombre de liens pointant vers le site mais aussi de la « popularité » des sites d'où proviennent ces liens. L'algorithme PageRank breveté par Google permet de classer les sites en fonction de leur impact sur le web. Il n'est pas sans rappeler l'impact factor utilisé pour le calcul de l'impact des revues scientifiques à partir du nombre de fois qu'elles sont citées en bibliographie.

Néanmoins, il ne s'agit pas que d'un simple audimat puisque d'autres critères secondaires déterminent le référencement des sites sur Google. Ainsi, un site ayant un nom de domaine gouvernemental ou possédant le

³ Pour une présentation plus générale de l'algorithme de Google, voir Grégoire G. (2012). *Le référencement sur Google, le guide complet*, Paris, Lavoisier.

mot recherché dans son nom de domaine a de forte chance d'être bien référencé.

La richesse sémantique et lexicale d'un site est également analysée. Enfin, le comportement lui-même des utilisateurs et le fait qu'ils restent suffisamment longtemps sur le site est aussi un indicateur de confiance (TrustRank). De la sorte, un site composé d'un contenu pauvre ou non original, c'est-à-dire issu dynamiquement d'autres sites web et sur lequel les usagers auraient tendance à ne pas rester, risque d'être mal référencé par Google et par la plupart des autres moteurs de recherches qui fonctionnent sur des algorithmes relativement similaires.

Améliorer le référencement des documents numérisés

Partant de ces considérations, il semble assez évident qu'une bibliothèque numérique qui proposera une riche variété de documents accessibles en texte intégral (c'est-à-dire numérisés) aura un PageRank supérieur à une bibliothèque numérique qui ne proposera qu'un nombre limité de documents accessibles en mode image. En effet, plus le nombre de différents documents accessibles via des permaliens sera important, plus le nombre de liens potentiels susceptibles de pointer vers cette bibliothèque numérique sera important et donc, plus le PageRank sera élevé. Cet état de fait va favoriser les bibliothèques numériques offrant un nombre important de documents. C'est en partie pour cette raison que la Bibliothèque Sainte-Geneviève a décidé de diffuser sa bibliothèque numérique sur la plateforme Internet Archive⁴ qui est l'une des plus considérables et des mieux référencées dans le monde. En effet, sur la base d'une étude comparative menée préalablement⁵, il est ressorti que cette plateforme est la 2^e dans le monde aussi bien en terme de nombre de livres numérisés en accès libre et gratuit qu'en terme de référencement sur le web. Ses fonctionnalités avancées (EPUB générés, flux RSS, lecture pour malvoyants, compatibilité Zotero, gestion des licences juridiques, moissonnage en OAI-PMH) ont également conforté le choix de participer à ce projet international à but non lucratif et membre de l'Open Content Alliance, puisqu'Internet Archive propose 3 millions de livres numérisés et s'affiche par conséquent comme l'un des seuls concurrents crédibles de Google. Avec ses 15 millions de dollars américains par an, ses 50 programmeurs, bibliothécaires et administrateurs et ses 150 scannéristes, mais aussi ses serveurs miroirs

⁴ <http://archive.org/details/bibliothequesaintegenevieve> [lien vérifié le 11 juillet 2012].

⁵ Andro M., Asselin E., Maisonneuve M. (2012). *Bibliothèques numériques : logiciels et plateformes*, Paris, ADBS.

déployés aux États-Unis, en Hollande, en Égypte et en Inde, elle dispose d'un budget qu'aucune bibliothèque universitaire française ne pourrait raisonnablement mobiliser.

Afin d'améliorer encore ce résultat et cette visibilité, il a été décidé dès le début du projet que chaque document numérisé et mis en ligne serait également signalé dans le catalogue national SUDOC⁶ et, par synchronisation, dans le catalogue en ligne de la bibliothèque. Cela contribue à augmenter le nombre de liens qui pointent vers la bibliothèque numérique et à en améliorer le référencement.

Enfin, Internet Archive dispose d'un entrepôt OAI pour chacune de ses collections. Les données relatives aux livres numérisés par la Bibliothèque Sainte-Geneviève apparaissent par exemple dans Gallica (bibliothèque numérique de la Bibliothèque nationale de France) ou dans Isidore (CNRS) grâce au moissonnage OAI-PMH⁷ : ce protocole d'échange permet d'afficher des métadonnées provenant de différents entrepôts sur des plateformes variées. Cela ajoute encore des liens vers la bibliothèque numérique, améliore le référencement et la visibilité des documents mis en ligne et pourra permettre à l'avenir à la Bibliothèque Sainte-Geneviève d'être sélectionnée dans le cadre d'appels à projets de la Bibliothèque nationale de France, le moissonnage OAI-PMH étant une condition requise.

En résumé, chaque document numérisé produit est désormais accessible depuis :

- Moteurs de recherche (les documents numérisés sont très bien référencés et apparaissent généralement en 1^{ère} position dans les pages de réponses de Google)
- Internet Archive
- Gallica (par moissonnage OAI-PMH)
- Europeana (par moissonnage OAI-PMH de Gallica)
- Isidore (par moissonnage OAI-PMH)
- SUDOC
- OPAC de la Bibliothèque Sainte-Geneviève

⁶ Système Universitaire de Documentation.

⁷ Open Archives Initiative Protocol for Metadata Harvesting.

Cette stratégie de mise en valeur des métadonnées a été adoptée par nombre de bibliothèques partenaires de la BnF, qu'il s'agisse de Medic@⁸, du Cnum⁹, de la Jubilotheque¹⁰ ou encore de la bibliothèque numérique de l'INHA¹¹. Si elles ont fait le choix d'une plateforme de diffusion propre, au contraire de la Bibliothèque Sainte-Geneviève, elles ont en revanche également prévu le moissonnage de leurs métadonnées par Gallica.

Bibliométrie de la numérisation

Les statistiques de consultation d'une bibliothèque numérique demeurent le principal moyen objectif de l'évaluer.

La Bibliothèque Sainte-Geneviève est la première bibliothèque en France à participer à la plateforme internationale à but non lucratif Internet Archive. Les statistiques sont obtenues via un module spécifique intégré à la plateforme. Elles ne permettent pas, pour le moment, de connaître origines géographiques, réseaux et mots clés, mais permettent au moins de connaître avec précision le nombre de téléchargements par document ou pour toute la collection et de croiser ces données avec la variable temps. Durant l'année 2011, les documents numérisés ont déjà été téléchargés 28 850 fois, avec une moyenne de 3,54 consultations par document et par mois. Ce résultat encourageant conforte le choix qui a été fait de la plateforme Internet Archive pour la diffusion.

La relative fréquentation de la bibliothèque numérique s'explique par les plus de 3 millions de documents proposés par Internet Archive. En effet, cette masse, bien supérieure à ce que propose l'ensemble des bibliothèques numériques françaises, génère un nombre important de liens vers Internet Archive qui dispose ainsi d'un bon indice PageRank de référencement, cet algorithme prenant largement en compte le nombre de liens pour mesurer l'impact et la popularité d'un site web. C'est cet indice qui détermine ensuite le classement des sites, dans la liste des résultats à une requête sur Google, et c'est, en grande partie, parce que cet indice est de 8/10 pour Internet Archive que la bibliothèque numérique de la Bibliothèque Sainte-Geneviève

⁸ Bibliothèque numérique de la Bibliothèque Interuniversitaire de Santé : <http://www.bium.univ-paris5.fr/histmed/medica.htm> [Lien vérifié le 12/07/2012].

⁹ Bibliothèque numérique du Cnam : <http://cnum.cnam.fr> [Lien vérifié le 12/07/2012].

¹⁰ Bibliothèque numérique de l'Université Pierre et Marie Curie : <http://jubilotheque.upmc.fr> [Lien vérifié le 12/07/2012].

¹¹ Bibliothèque numérique de l'Institut national d'histoire de l'art : <http://bibliotheque-numerique.inha.fr> [Lien vérifié le 12/07/2012].

est visible sur le web et que les livres numérisés obtiennent une forte fréquentation.

Convaincue qu'un bon référencement était lié à une masse importante de documents numérisés à proposer et que cette masse ne pouvait être que le résultat d'une mutualisation de la diffusion, la Bibliothèque Sainte-Geneviève a proposé au PRES Sorbonne Paris-Cité et à la Ville de Paris le développement d'une plateforme collective.

Le projet : www.bibliotheque-numerique.fr¹²

Une partie importante des documents numérisés par les bibliothèques en France n'est pas diffusée en ligne et demeure sur des DVD dont la durée de vie est limitée à moins de 10 ans. En effet, les bibliothèques renoncent bien souvent à développer des plates-formes de diffusion qui leur seraient propres car trop coûteuses en ressources humaines et financières pour un résultat souvent décevant.

Afin d'offrir enfin un débouché à ces numérisations, la Bibliothèque Sainte-Geneviève a proposé au PRES¹³ Sorbonne Paris-Cité un projet de plateforme de diffusion mutualisée. Celle-ci permettra de partager les coûts de développement, d'administration et de maintenance, d'accroître la visibilité des projets de numérisation grâce à un volume cumulé de documents numériques et de limiter leur éparpillement, mais aussi, d'en améliorer la qualité et d'en garantir la pérennité. Malgré le caractère collectif de la plateforme, l'identité de chaque institution pourra être maintenue (nom de domaine, graphisme de la vitrine, logo, statistiques de consultation).

Cette plateforme se distinguera de celles qui existent déjà par des fonctionnalités innovantes.

La numérisation à la demande

À partir des catalogues en ligne des bibliothèques et à partir de la plateforme collective, le financement de la numérisation de chaque livre libre de droit sera proposée aux internautes, aux institutions et aux mécènes qui pourront en commander la numérisation via des boutons dynamiques renvoyant vers un paiement en ligne sur le modèle du réseau européen Ebooks on Demand¹⁴. Dans le cadre d'une délégation de service public, un

¹² Andro M. (2011). Genèse d'un projet de plateforme mutualisée pour la diffusion des documents numérisés, *Arabesque*, n° 64.

¹³ Pôle de Recherche de l'Enseignement Supérieur.

¹⁴ <http://books2ebooks.eu/fr> [Lien vérifié le 12/07/2012]. La BIU Santé a été la première bibliothèque française à intégrer ce réseau innovant de service.

prestataire de numérisation viendra périodiquement, par navette, rechercher et numériser les documents ainsi commandés.

La mention « ce livre a été numérisé grâce au soutien de tel particulier, de telle équipe de recherche, de telle fondation ou de tel investisseur », sur le modèle des publicités Google Adwords, permettra de favoriser et d'encourager ce financement participatif, y compris sous la forme de crowdfunding. Pour leur part, les bibliothèques pourront offrir à leurs usagers des services de reprographie numérique, encore relativement peu proposés par les bibliothèques, sans avoir à en supporter le coût ni à en assurer la gestion. Elles auront aussi la possibilité de compléter leur programmes de numérisation tout en partageant leur politique de sélection des documents à numériser avec le grand public, des investisseurs et des mécènes.

L'impression à la demande

Dans le secteur de l'édition, le modèle économique de l'impression à la demande est en plein développement. Il s'agit d'imprimer en flux tendu, en fonction de la demande et d'éviter les invendus ainsi que les coûts induits par la gestion des stocks.

Les documents numérisés par les bibliothèques pourront être vendus, par correspondance, par une société délégataire de service public, depuis sa propre plateforme de paiement et depuis la plateforme collective des bibliothèques, sous la forme de livres imprimés et brochés, à l'image de ce que la bibliothèque de l'Université du Michigan a mis en place en partenariat avec Amazon Booksurge.

Une ou plusieurs Espresso Book Machine pourront également être installées dans des bibliothèques à forte fréquentation. Ces machines offriront la possibilité au public de commander sur place l'impression papier et brochée de près de 7 millions de livres libres de droits numérisés par Google Books, Internet Archive et même d'une partie des documents numérisés par la Bibliothèque nationale de France et diffusés sur Gallica.

La numérisation comme l'impression à la demande ne se limitent pas à la perspective gestionnaire d'un retour sur investissement. Il s'agit prioritairement d'offrir des services nouveaux aux usagers en faisant supporter leurs coûts au privé. Ce modèle économique devrait être profitable à la fois aux usagers disposant de nouveaux services (reprographie numérique, impressions brochées de livres numérisés), aux bibliothèques (retour sur investissements), aux mécènes (possibilité de soutenir des projets de numérisation), aux investisseurs (retour sur investissements sous forme d'affichage de sites et de trafic web en échange

du financement d'opérations de numérisation), et pour finir, aux prestataires (création d'activité de numérisation et d'impression).

La correction participative de l'OCR¹⁵

Les textes numérisés tels qu'ils résultent de processus de reconnaissance automatique de caractères (OCR) présentent de nombreuses erreurs liées à des typographies anciennes ou singulières, à des artefacts, des annotations... Ces erreurs de reconnaissance de caractères sont traditionnellement corrigées par les prestataires de numérisation qui ont le plus souvent recours à de la main d'oeuvre à bas coûts dans des pays en voie de développement comme Madagascar, le Viêt Nam ou l'Inde.

La plateforme bibliotheque-numerique.fr proposera une correction participative du texte ocrisé, selon les principes du crowdsourcing¹⁶, et sur le modèle de l'Australian Newspapers Digitisation Program ou de Wikisource (Wikipedia¹⁷).

L'encodage XML TEI (Text Encoding Initiative)

À partir du texte ocrisé, les bibliothécaires pourront encoder par des balises XML des phrases ou des mots afin de structurer le texte et d'en permettre une valorisation scientifique et éditoriale en générant des index de noms de personnes, de noms de lieux, de noms de métiers *etc.* sur le modèle de ce qui est réalisé par les Bibliothèques Virtuelles Humanistes.

Cette activité de description bibliographique à l'intérieur des textes représente, pour une part, l'avenir de la profession de bibliothécaire. Elle rend possible une exploitation scientifique, statistique et bibliométrique des corpus numérisés. Par exemple, les historiens pourront, grâce à elle, générer des représentations de réseaux sociaux ou des cartes géographiques d'échanges.

L'annotation collaborative

¹⁵ Optical Character Recognition.

¹⁶ Pour en savoir plus sur le crowdsourcing appliqué aux bibliothèques : Holley R. (2010). Crowdsourcing: How and Why Should Libraries Do It?, *D-Lib Magazine*, n°6.

¹⁷ Après avoir expérimenté la correction participative de l'OCR de thèses de l'École Nationale Vétérinaire de Toulouse, Wikimedia a d'ailleurs passé un accord avec la BnF en 2010 pour permettre aux internautes de participer à la correction de l'OCR d'environ 1 400 documents présents sur Gallica : <http://blog.wikimedia.fr/tag/ocr> [Lien vérifié le 12 juillet 2012]. Voir également une interview du président de Wikimedia France à ce sujet : <http://www.bibliobsession.net/2010/04/13/accord-bnf-wikimedia-nous-preparons-le-web-de-donnees> [Lien vérifié le 12 juillet 2012].

À l'image de ce qui est pratiqué par le Centre de ressources et d'information multimedia pour l'enseignement supérieur (Cerimes) avec le projet bibnum¹⁸, la plateforme offrira la possibilité d'accéder à des présentations, des analyses, des synthèses, des notes... sur les documents numérisés. Ce sera l'occasion, pour les bibliothécaires de revenir à leur coeur de métier qui n'est ni le contrôle qualité des documents numérisés, ni même leur mise en ligne.

Cette plateforme de bibliothèque numérique mutualisée et innovante a vocation à être ouverte bien au-delà des seules bibliothèques universitaires du PRES Sorbonne Paris-Cité et à offrir, en particulier, un débouché à toutes les bibliothèques de la ville de Paris et de la région Ile-de-France et probablement encore au delà. Cette plateforme pourrait en quelque sorte et, toutes proportions gardées, devenir l'Internet Archive français. Dans l'attente de l'avancement de ce projet, certes voté et financé mais conservant, comme tout projet, une part d'incertitude, la Bibliothèque Sainte-Geneviève a néanmoins souhaité communiquer et valoriser le fruit de sa numérisation : le site web est un des éléments supplémentaires de cette démarche.

Développement d'une vitrine sur le site web de la Bibliothèque Sainte-Geneviève

La participation à un projet coopératif international pouvant parfois être vécue comme une perte d'identité, il était nécessaire, aussi bien pour la communication interne (personnels) que pour la communication externe en direction des usagers actuels et des autres bibliothèques, de développer une vitrine spécifique sur le site web de la bibliothèque.

Des pages web ont été développées afin de présenter les différents projets de numérisation de la Bibliothèque.

¹⁸ <http://bibnum.education.fr>

BIBLIOTHÈQUE SAINT-GENEVIÈVE
 Bibliothèque Sainte-Geneviève
 La BSG • Infos Pratiques • Collections • Catalogues • Ressources • Services • Amis de la BSG • Actus

Forum
 Plan du site
 Aide
 Actus

Bibliothèque numérique

Présentation Incunables Voyages nordiques Livres rares du 19e

Le Catalogue Informatisé
 Le Fonds Général
 La Réserve
 La Nordique
 Henri Labrousse

Environ 300 incunables et post-incunables ont été sélectionnés en raison de leur rareté, de leur singularité en matière de variantes typographiques, de leur décor ou de la présence d'annotations manuscrites remarquables. Ainsi, la plupart des exemplaires uniques (unica) de la collection ont été numérisés. En parallèle, la Bibliothèque Sainte-Geneviève a accéléré le catalogue de ses incunables.

[Bible. Latin. 1462] Biblia latina. — Mainz : Johann Fust et Peter Schöffer, [14 VIII] 1462. — 2 part., 2°.	Document numérisé
[Bible. Latin. 1476.] — Interpretationes Hebraicorum nominum. — Venezia : Nicolas Jenson, 1476. — 2°.	Document numérisé
[Bible. N. T. Français.] Cy commence le nouveau testament. Ed. Julien Macho et Pierre Farget. — Lyon : [Guillaume Le Roy pour] Barthélemy Buyer, [circa 1476-1478]. — 2°.	Document numérisé
[Biblia. Ital.]. — [Venezia : Adam de Ambergau,] 1 x 1471. — 2 part., 2°.	Document numérisé
[Biblia. Latin.]. — Mainz : Peter Schöffer, [23 ii] 1472. — 2 part., 2°.	Document numérisé
[Breviarium Bituricense.]. — Venezia : [Adam de Rotweil], Petrus de Piasis de Cremona, Bartholomaeus de Blavis de Alexandria et Andreas Torresanus de Adula pour Guillermi Touze, 23 I 1481. — 8°.	Numérisation en cours
[Breviarium parisiense] Éd. Jean Le Munerat. — Paris: Jean Du Pré, 9 vi 1492. — 7 parties, 2°, ill.	Document numérisé
[Breviarium Romanum]. — Torino : Johannes Fabri et Johanninus de Petro pour Pantaleone da Confienza, 1474. — 8°.	Document numérisé
[Calendrier des bergers.] Kalendar et compost des bergiers. — Paris : Gaspard Philippe [inter 20 viii 1506 et 1510]. — In-4°, ill.	Document numérisé

On peut signaler entre autres un exemplaire d'une édition latine de l'Histoire des animaux d'Aristote ayant appartenu au conseiller du roi Jean Budé et annoté par son fils, le célèbre humaniste et philologue Guillaume Budé (cote : OEXV 181 RES)

Historia Search inside

retour haut de page

Figure 1. Page web de présentation des incunables numérisés

Chaque document numérisé et diffusé sur Internet Archive peut être visualisé directement sur une page du site web de la Bibliothèque. Il est également possible d'y ajouter notices bibliographiques complètes, commentaires, analyses, présentations...

concurrentiel du web, maintenir une bonne visibilité des collections ainsi promises à un public plus large passe aussi par une attention soutenue portée à la question du référencement et des services associés. Cette question devrait d'ailleurs inciter les bibliothèques à davantage de mutualisation si elles souhaitent améliorer la visibilité de leurs programmes de numérisation sur le web.

La valorisation des bibliothèques numériques impose également un dispositif de veille continue. À la Bibliothèque Sainte-Geneviève, des flux RSS¹⁹ sur des revues en sciences de l'information, des bases bibliographiques, des blogs, des sites institutionnels, des sites de presse permettent ainsi d'identifier des opportunités et des innovations technologiques (encodage TEI, correction participative du texte OCRisé, numérisation à la demande, vente d'impressions à la demande...) ou stratégiques (partenariats, appels à projets...) afin de garantir l'adaptation de la bibliothèque numérique et des services proposés aux besoins émergents des usagers.

Durant la phase de conduite de projet, la perspective technologique ne doit pas être opposée à la perspective gestionnaire, pas plus que les préoccupations des bibliothèques ne doivent l'être avec celles de leurs usagers. Dans le domaine de la numérisation en particulier, des harmonies peuvent être trouvées au bénéfice des usagers, des bibliothèques et des entreprises. C'est le cas, en particulier de la numérisation et de l'impression à la demande susceptibles d'offrir des services nouveaux aux usagers, génèrent des retours sur investissements aux bibliothèques, créent de l'activité pour des prestataires, apportent de la visibilité pour des mécènes et engendrent du trafic web pour des investisseurs dans la numérisation de documents.

Bibliographie

Andro M., Asselin E., Maisonneuve M. (2012). *Bibliothèques numériques : logiciels et plateformes*, Paris, ADBS.

Chamberlaina E. (2012). Investigating Faster Techniques for Digitization and Print-on-Demand, *New Review of Academic Librarianship*, n° 18.

Dalmau M., Hawkins K. (2010). *Best Practices for TEI in Libraries*, <http://purl.oclc.org/NET/teiinlibraries> [Consulté le 12 juillet 2012].

Grégoire G. (2012). *Le référencement sur Google, le guide complet*, Paris, Lavoisier.

¹⁹ Rich Site Summary.

Mathis R. (2011). Déficiences et production, *BBF*, n°1, p. 10-13, <http://bbf.enssib.fr> [Consulté le 12 juillet 2012].

Miller M., Choi G., Chell L. (2012). Comparison of three digital library interfaces : open library, Google books, and Hathi Trust, *Proceedings of the 12th ACM/IEEE-CS joint conference on Digital Libraries*.

Mülberger G., Gstrein S. (2009). eBooks on Demand (EOD) : a European digitization service, *IFLA Journal*, n° 35.

Vandewalle J.-J. (2011). Signalement, numérisation et valorisation, *BBF*, n°2, p. 91-93, <http://bbf.enssib.fr> [Consulté le 12 juillet 2012].

Wilson-Higgins S. (2011). Could print on-demand actually be the « new interlibrary loan »?, *Interlending & Document Supply*, n°39.